**Akhil Pathak**

  Texas, Dallas 75071 ♦ 9803586093 ♦ akhilpathak1234@gmail.com

Professional Summary

Result-oriented Data Engineer with expertise in designing advanced, cost-efficient Big Data ETL/ELT solutions in the cloud, complete with scalability, monitoring, logging, and alert systems. Expert in crafting cloud-native projects with hands-on experience in Snowflake and Amazon (AWS). Collaboratively engaged with data analysts and machine learning teams to elevate data-driven solutions. Possesses a good understanding of cybersecurity fundamental like threat hunting, Vulnerability management and hands on experience in exploring various tools like Splunk, Cribl Cloud and AWS Security Lake.

Skills

* Cloud - **Snowflake, AWS, Azure, GCP, Cribl**
* AWS - **Redshift, Lambda, S3, EMR, Glue, SQS**
* Machine Learning **- Pandas, NumPy, sklearn**
* Programming Languages - **Python, SQL, Shell scripting, JavaScript, React**, **NodeJS**
* ETL/ELT - **DataBricks, Matillion**
* Databases - **Postgres, MySQL, SQL Server, Oracle**
* Framework- **Spark, PySpark**
* Version Control - **GIT, SVN**
* NoSQL Databases - **MongoDB, Redis**
* AI**: Rags, Hugging Face Transformers, OpenAI API, AWS Bedrock, Llama 3, Claude**

Work History

**Python AWS Data Engineer,** 11/2023 to 06/05/2025

**CGI Inc -** Texas, Plano,Fannie Mae

* Collaborated with cross‑functional teams to build a centralized checklist that evaluates control profiles, generates System Security Plan (**SSP**) PDFs, and assigns ownership for protocol implementation before release.
* Built a high-throughput ETL pipeline to process **300K+ S3** records per run, transform metadata, and expose enriched data via REST API, accelerating access for downstream engineering teams.
* Automated SSP document generation using validated control-profile data, cutting down manual prep time by **50%** and enabling faster approvals from governance teams.
* Integrated **600K+** CMDB and vulnerability records into an LLM-powered **SQL Decoder** using LangChain in Python, automating **~500** monthly queries and reducing security analyst response time by **25%** during incident triage.
* Built a responsive chatbot UI in React, connected to a Flask API for real-time interactions—boosted engagement by **20%.**
* Deployed Flask backend via Docker on AWS ECS behind an ALB, introduced auto-scaling to support **250+** concurrent users with **99.95%** uptime and **25%** faster container spin-ups.
* Led a **30-day** POC to compare Cribl Cloud vs. Splunk for **1 TB/day** log ingestion, achieved **35%** licensing cost savings, enabling more affordable log analytics at scale for the security operations team.
* Developed Cribl pipelines to ingest and normalize **AWS VPC Flow, Zscaler, and SentinelOne logs**, enabling security analysts to query multi-petabyte datasets consistently and detect anomalies faster.
* Simulated realistic attack scenarios using Python’s Faker to generate synthetic logs, enabling robust testing of Cribl ingestion workflows without risking production data.
* Rebuilt and deployed a Docker-based GenAI chatbot on AWS ECS with auto-scaling and load balancing, delivering secure, low-latency access for over **1,000** internal users.

**Software/Data Engineer Intern**, 09/2023 to 10/2023

**Wise Equation Solutions**– Austin, Texas

* Developed Python and SQL scripts to process and prepare over **200,000** financial records daily for quarterly profit and loss dashboards, improving report accuracy by **15%.**
* **Oversaw the daily S3-to-Snowflake data pipeline, ensuring each day's data was loaded on time; this consistency reduced data latency by 20%.**
* **Synced daily with the finance team to double-check data integrity and ensur**e that da**shboard metrics matched their records. This teamwork provided leadership with accurate data for timely decisions.**
* Created an alert system for the data pipelines which typically resolved any issues within **30 minutes** to maintain **99.9%** uptime.

**Software Engineer**, 09/2017 to 07/2019

**LTI - Larsen & Toubro Infotech** – Mumbai, India

* Helped in optimizing data storage and retrieval using indexing and compression techniques, resulting in a **30%** improvement in query performance.
* Coordinated with the team to develop the best practices and coding standards, resulting in a **50%** reduction in code review time from **1 hour to 30 minutes**.
* Wrote down detailed Release documents and milestone updates to project managers and stakeholders, resulting in a **25%** decrease in deployment time.
* Coordinated with production team for ongoing production releases and responded to problems by diagnosing and correcting logic and coding errors resulting **Zero downtime**.
* Engineered conversion modules to seamlessly transform XML data to CSV and JSON to CSV, enhancing data interoperability and streamlining integrations.
* Evaluated project requirements and specifications and developed software applications that surpassed client expectations.
* Orchestrated efficient large-scale software deployments efficiently.
* Participated in entire software development cycle from start-up specifications to rollout of final product.

**Software Development & Research Intern**, 05/2015 to 08/2015

**NFC - Nuclear Fuel Complex** – Hyderabad, India

* Delivered solutions for Escalation Management of Workflow Systems which is an automated management system to keep track of biometric data of employees in Nuclear Fuel Complex.
* Developed Automated processes to automate manual tasks by using monthly recurring bash scripts resulting in **200x** increase in efficiency.
* Maintained Data Pipeline up-time of **99.8%** while ingesting transactional data across different primary data sources using tools like PostgreSQL, Java, Linux, and SQL.
* Generated a company-wide distribution list comprising of Analytics of Employee Attendance using Bio-Metric Data.

Education

**Auburn University**, **Montgomery** Master’s in Computer Science GPA:4.00/4.00 May 2023

**SRM University, Chennai** Bachelor’s in Computer Science GPA:3.58/4.00 May 2017